

Least Squares Estimation Techniques for Position Tracking of Radioactive Sources*

James W. Howse

Modeling, Algorithms and Infomatics Group, CCS-3, Mail Stop F645
Los Alamos National Laboratory
Los Alamos, NM 87545

Lawrence O. Ticknor

Statistics Group TSA-1, Mail Stop F600
Los Alamos National Laboratory
Los Alamos, NM 87545

Kenneth R. Muske

Department of Chemical Engineering
Villanova University
Villanova, PA 19085

Abstract

This paper describes least squares estimation algorithms used for tracking the physical location of radioactive sources in real-time as they are moved around in a facility. We present both recursive and moving horizon nonlinear least squares estimation algorithms that consider both the change in the source location and the deviation between measurements and model predictions. The measurements used to estimate position consist of four count rates reported by four different gamma ray detectors. There is an uncertainty in the source location due to the large variance of the detected count rate, and the uncertainty in the background count rate. This work represents part of a suite of tools which will partially automate security and safety assessments, allow some assessments to be done remotely, and provide additional sensor modalities with which to make assessments.

1 Introduction

In this paper we describe recursive and moving horizon algorithms for recursively estimating the real-time positions of radioactive sources in a facility. In this work, we will explicitly consider the problem of tracking a single source within one room. We estimate the position by assuming a known initial source position and then estimating the change in its position from this initial state. Our data is four time series consisting of the count rate at one second intervals from four gamma ray detectors which are located at four different positions within the room. The count rate at a particular sensor is the total number of gamma-energy photons received by the sensor during

*This work was performed under the auspices of the Department of Energy under contract W-7405-ENG-36.

a one second time interval. The specific sensors that we use consist of a photoluminescent slab of plastic attached to a photomultiplier tube. These sensors *do not* form an image of the room in the way that a camera would. This means that many of the techniques for locating moving objects in images can not be applied. Our approximations of the detected count rate are based on a nonlinear model of each sensor which relates the source location to the measured count rate at that sensor.

Our algorithm is a recursive nonlinear least squares estimator which simultaneously minimizes the expected value of the difference between the detector measurements and the model predicted count rates, and also the expected value of the change in the source location, in the presence of noise. The predominant source of noise comes from the stochastic nature of gamma emissions from the source itself. Ideally, the emission of gamma photons from a radioactive source over time looks like a series of samples drawn from a Poisson distribution. Since the variance of a Poisson distribution is equal to its mean, the deviation of the detector readings is proportional to the square root of the mean count rate. Hence the uncertainty in the source location based on the detector readings is approximately proportional to the square root of the mean count rate.

An analogy may clarify the difficulties associated with this problem. Imagine a building containing only one room and having a flat roof with four skylights cut into the ceiling, each near one corner of the building. You are standing on the roof and someone is walking around in the room holding a candle. You must determine the position of the candle in the room by observing the relative brightness of the light coming through the four skylights. Keep in mind that the flickering of the candle leads to variations in its brightness that are proportional to the square root of the brightness itself. This analogy makes it clear that many image processing techniques for tracking moving objects would be ineffective on this problem.

The facility in question conducts various experiments using radioactive materials. Knowledge of the source locations has implications for both security and safety. Currently all security and safety assessments are made by people who are physically present during experiments. The overall purpose of the project is to provide tools to facilitate and enhance this assessment process. Our algorithm is one part of a suite of tools which will allow some of these assessments to be made remotely. The tools will also provide sensor modalities that are not available to people using only their own senses. These tools will also be used to partially automate the assessment process. These three factors will enhance both security and safety by reducing personnel risk through remote assessment, providing new methods and sensor modalities for risk assessment, and providing a source of independent verification for the current assessment process.

2 Sensor Model

In this section we discuss the model that relates the detected count rate from a point source to the position of that source based on the analysis in Tsoulfanidis (1983). For the i th detector, the relationship between the detected count rate \mathcal{D}_i and the source position (u_i, v_i, w_i) relative to that detector is given by

$$\begin{aligned} \mathcal{D}_i &= \frac{\Omega_i(u_i, v_i, w_i) \mathcal{S} \epsilon_i \mathcal{F}_i + \mathcal{B}_i}{1 + \tau_i \Omega_i(u_i, v_i, w_i) \mathcal{S} \epsilon_i \mathcal{F}_i} \\ &\Rightarrow \mathcal{D}_i = \mathcal{M}_i(u_i, v_i, w_i) \end{aligned} \tag{1}$$

for $i = 1, 2, \dots, m$, where in our case $m = 4$. In this equation \mathcal{S} is the actual source strength, ϵ_i is the detector efficiency, τ_i is the dead time, \mathcal{F}_i is the product of all the correction factors (e.g., absorption and backscattering), \mathcal{D}_i is the total number of counts per unit time actually detected, \mathcal{B}_i is the number of counts per unit time which constitute the background, and the view factor $\Omega_i(\cdot)$ is the ratio of the number of particles which actually enter the detector to the total number of particles emitted by the source. The quantity $\Omega_i(\cdot)$ can also be thought of as the solid angle subtended by the detector for a particular source location, which ranges between 0 and 4π . For these detectors, the efficiency is rated at $\sim 10\%$ for a Cs^{137} source. The *dead time* of a detector is defined to be the minimum length of time which must separate incident photons in order for them to be recorded as two separate pulses. The photoluminescent material used in these detectors has a dead time of 3.3 nanoseconds. Note that there is one equation of this form for each of the m detectors. We will use the right hand side of Equation (1) as an approximation for the count rate given a particular position (u_i, v_i, w_i) , and we denote this function by $\mathcal{M}_i(\cdot)$.

For a point source and a rectangular detector of finite size, the most general relative position is shown in Figure 1. The solid angle $\Omega_i(\cdot)$ subtended by a detector of width \mathcal{W} and height \mathcal{H} for

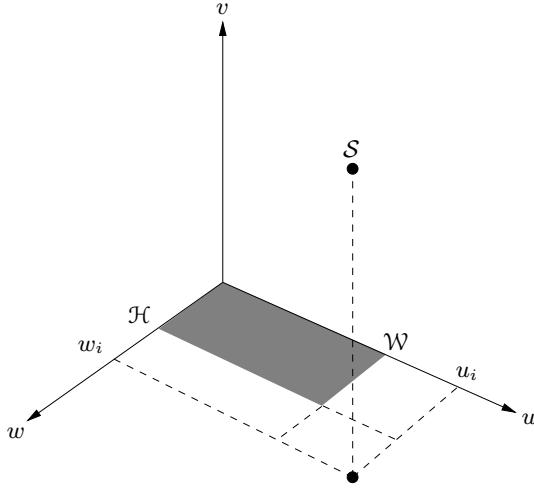


Figure 1: The relative position between a point source at position \mathcal{S} and a rectangular detector of width \mathcal{W} and height \mathcal{H} . The projections of the source location onto the u and w axes are u_i and w_i respectively.

a point source located at (u_i, v_i, w_i) , relative to the i th detector, is given by

$$\begin{aligned} \Omega_i(u_i, v_i, w_i) &= \arctan \left(\frac{u_i w_i}{|v_i| \sqrt{u_i^2 + w_i^2 + v_i^2}} \right) - \arctan \left(\frac{(u_i - \mathcal{W}) w_i}{|v_i| \sqrt{(u_i - \mathcal{W})^2 + w_i^2 + v_i^2}} \right) \\ &\quad - \arctan \left(\frac{u_i (w_i - \mathcal{H})}{|v_i| \sqrt{u_i^2 + (w_i - \mathcal{H})^2 + v_i^2}} \right) + \arctan \left(\frac{(u_i - \mathcal{W}) (w_i - \mathcal{H})}{|v_i| \sqrt{(u_i - \mathcal{W})^2 + (w_i - \mathcal{H})^2 + v_i^2}} \right) \end{aligned} \quad (2)$$

as shown in Gotoh and Yagi (1971). The detectors that were used have a height \mathcal{H} of 0.83 feet and a width \mathcal{W} of 3.00 feet. The minimum value of $\Omega_i(u, v, w)$ is 0, and its maximum value is 2π since even when the source is touching a planar detector, only half of the emitted photons will actually strike the detector. The variation of the solid angle $\Omega_i(u, v, w)$ with position is illustrated in Figure 2. Figure 2(a) shows the variation in the u - v plane with the source fixed along the w -axis at $\frac{\mathcal{H}}{2}$, and Figure 2(b) plots the change in the w - v plane with the source fixed along u at $\frac{\mathcal{W}}{2}$. Note

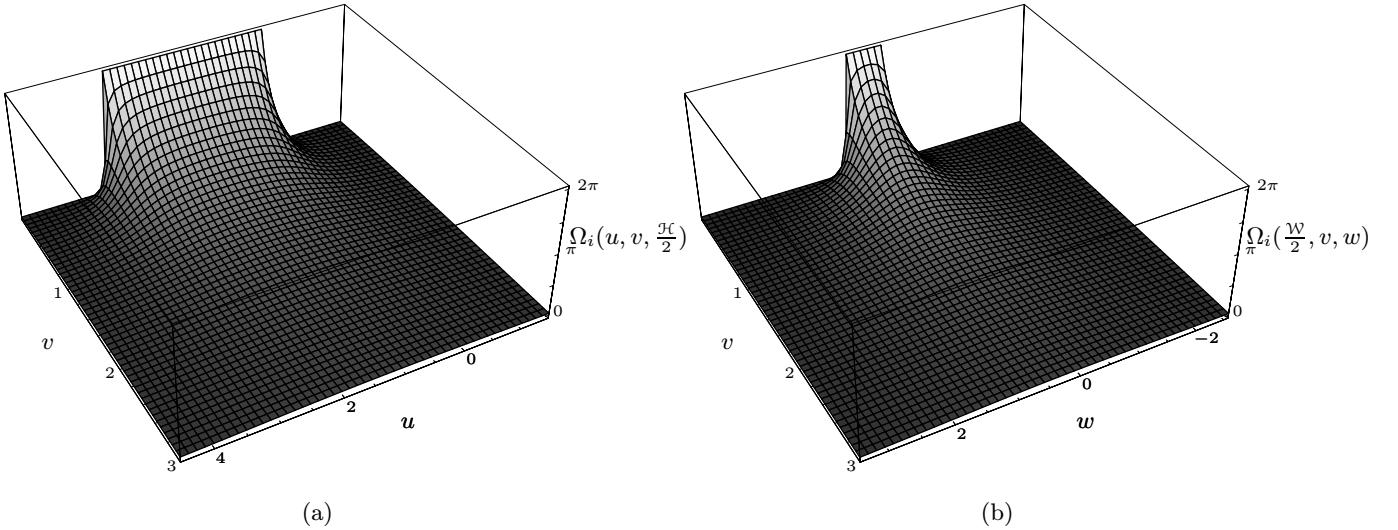


Figure 2: The variation of the solid angle $\Omega_i(u, v, w)$ in the (a) u - v plane (b) w - v plane. The third dimension was fixed at (a) $w = \frac{\mathcal{L}}{2}$ (b) $u = \frac{\mathcal{L}}{2}$.

that the solid angle is always positive and approaches 2π as the source gets closer to the detector. Also the solid angle decreases monotonically both as the source-detector distance increases and as the source-detector angle increases. We define source-detector angle as the angle measured relative to the line perpendicular to the center of the detector. Hence this angle is 0 when the source is directly above the center of the detector.

It is clear from Figure 1 that the coordinate system used to derive Equation (2) is detector centered, *not* room centered. Since the desired answer is the source location in room centered coordinates, a transformation must be made between these two coordinate systems. The required transformation converts room centered coordinates (x, y, z) into detector centered coordinates (u_i, v_i, w_i) , which allows Equation (2) to be properly evaluated. One way to construct this transformation is by moving each of the detectors from their actual location to the selected origin of the room. This is accomplished by first rotating the sensor about the w -axis in Figure 1, and then translating the rotated sensor to the origin of the room. This transformation can be written in matrix form as

$$\begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} = \begin{pmatrix} \cos \mathcal{A}_i & \sin \mathcal{A}_i & 0 & \mathcal{T}_{x,i} \cos \mathcal{A}_i + \mathcal{T}_{y,i} \sin \mathcal{A}_i \\ -\sin \mathcal{A}_i & \cos \mathcal{A}_i & 0 & \mathcal{T}_{y,i} \cos \mathcal{A}_i - \mathcal{T}_{x,i} \sin \mathcal{A}_i \\ 0 & 0 & 1 & \mathcal{T}_{z,i} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (3)$$

where $\mathcal{T}_{x,i}$, $\mathcal{T}_{y,i}$, and $\mathcal{T}_{z,i}$ are the translations along the room centered x , y , and z directions, respectively, for the i th detector. The quantity \mathcal{A}_i is the rotation angle around the w -axis for the i th detector. Note that the source location in room centered coordinates is (x, y, z) . Also note that the source strength \mathcal{S} is independent of position, so it is unchanged under coordinate transformation.

3 Estimation Algorithm

In this section we discuss the estimation algorithm used to track the position of the radioactive source over time. We begin by discussing how to use the model in Equation (1) to locate a radioactive source. Assume for the moment that the detected count rate \mathcal{D}_i obeys the deterministic model in Equation (1). For a single source, the count rate \mathcal{D}_i at the i th detector defines a single level curve of $\mathcal{M}_i(u, v, w)$ given by $\mathcal{M}_i(u, v, w) = \mathcal{D}_i$. There is one such curve for each detector and their intersection points define the possible source locations, as shown in Figure 3 where the level surfaces are projected into the x - y plane. The perspective of this figure is looking down into

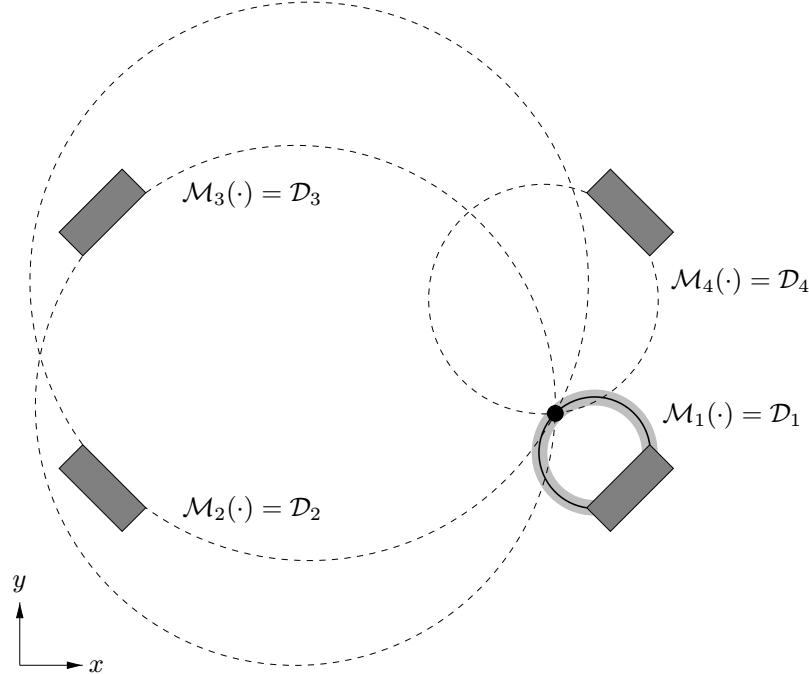


Figure 3: The intersection of a set of level surfaces for the four detectors projected into the x - y plane of the room.

the x - y plane of the room. The four dark grey boxes are the gamma detectors. The solid line and the three dotted lines each represent a level surface for a detector and the black dot • shows their intersection point. Note that the actual level surfaces are approximately ellipsoids, except for a neighborhood near the edges of the detector, but their projections into the x - y plane are drawn as circles in Figure 3 for convenience. Since the count rate is assumed to be consistent with the model, these four surfaces intersect at only one point, which corresponds to the source location. Now assume that the count rate \mathcal{D}_i is randomly distributed with a mean \mathcal{M} given by the model in Equation (1) and a variance \mathcal{V} proportional to the mean. In this case the level surfaces become shells whose thickness is proportional to $\sqrt{\mathcal{V}}$ as illustrated by the light grey band surrounding the solid black line in Figure 3. The intersection of these shells becomes some volume element in 3-dimensions wherein the source is located. So in this stochastic setting, the problem is to find the most likely location for a given set of detected count rates.

This problem was posed as a state estimation problem whose formulation is explained, for instance, in Sage and Melsa (1971), Jazwinski (1970), and Bryson and Ho (1975). In this context the measurements or outputs are the detected count rate $\mathcal{D}_i(k)$ at each detector i at each time step k . The states are chosen according to the problem that must be solved. In this case, the

state must at least contain the source location $(x(k), y(k), z(k))$ at every time step k . In addition to the source location, another unknown in Equation (1) is the background count rate \mathcal{B}_i at each detector. We assume that the background is the same at each detector, making $\mathcal{B}(k)$ independent of i , and include this single quantity in the state. Denote any change in the state estimates at time k by the vector $\mathbf{d}(k) = [\delta x(k) \delta y(k) \delta z(k) \delta \mathcal{B}(k)]^\top$, denote the state estimates at a particular time step k by the vector $\mathbf{e}(k) = [x(k) y(k) z(k) \mathcal{B}(k)]^\top$ and call the number of elements in each of these vectors $p = 4$. We chose to approximate the detected count rates with the vector $\hat{\mathbf{D}}(k)$ drawn from a Poisson process with mean vector $\mathcal{M}(\mathbf{e}(k))$, where $\mathbf{e}(k)$ is the current state estimate. The structure for our approximate system is shown in Figure 4. In this figure, $\mathcal{M}(\cdot)$ is the vector form

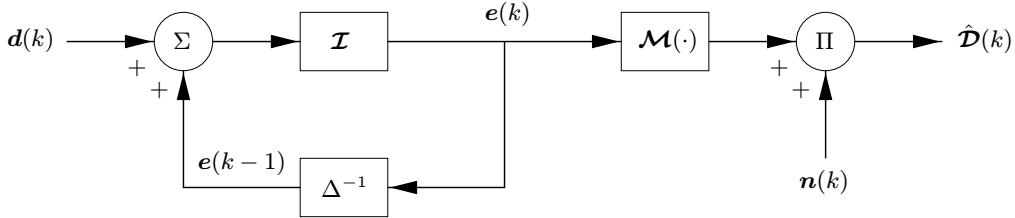


Figure 4: A block diagram of the system approximation.

for the right hand side of Equation (1), \mathcal{I} is the identity mapping, Δ^{-1} is a unit time delay, and $\mathbf{n}(k)$ is a multiplicative noise process. Multiplicative noise is chosen because the *link function* defined in McCullagh and Nelder (1989) for a Poisson distribution is $\exp(\cdot)$. The link function is the monotone differentiable transformation under which the transformed measurements can be written as a model with additive noise. Under the link function $\exp(\cdot)$, the approximate count rate is $\hat{D}_i(k) = \exp(\tilde{\mathcal{M}}_i(\mathbf{e}(k))) \exp(\tilde{n}_i(k))$ which implies that $\ln(\hat{D}_i(k)) = \tilde{\mathcal{M}}_i(\mathbf{e}(k)) + \tilde{n}_i(k)$. In our case, define the deterministic part of the model by $\exp(\tilde{\mathcal{M}}_i(\mathbf{e}(k))) = \mathcal{M}_i(\mathbf{e}(k))$, and the stochastic portion by $\exp(\tilde{n}_i(k)) = n_i(k)$. Let the mean of the stochastic portion be $\mathcal{E}\{n_i(k)\} = 1$. Therefore the modeled mean count rate is $\mathcal{E}\{\hat{D}_i(k)\} = \mathcal{M}_i(\mathbf{e}(k))$. Since we want $\hat{D}_i(k)$ to have the properties of a Poisson process, its variance must also be $\mathcal{M}_i(\mathbf{e}(k))$. This will be the case if the stochastic process is chosen such that $\mathcal{E}\{n_i(k)^2\} = 1 + \frac{1}{\mathcal{M}_i(\mathbf{e}(k))}$. This means that the characteristics of the stochastic portion of the model are *dependent* on the deterministic part of the model, in marked contrast to typical system models with Gaussian noise.

Note that there are upper and lower bounds on the state change $\mathbf{d}(k)$ and on the state estimate $\mathbf{e}(k)$. The constraints on the estimate $\mathbf{e}(k)$ are imposed so that the estimated position can not be outside the room, and the estimated background can not exceed historical bounds for the background levels in the room. The constraints on the change in estimate $\mathbf{d}(k)$ are imposed in order to keep noise from causing the position to vary excessively from time step to time step. Conceptually these constraints are reasonable because the sources are moved by people, and there is a limit to how far a person can move in a given time interval.

We find the maximum a posteriori (MAP) estimate for the state at each time step. As pointed out in Sage and Melsa (1971), for a Gaussian noise process this is equivalent to finding a weighted least squares fit to the measurements. This turns out to be true even when the system is nonlinear. For Gaussian noise the proper weights are the noise covariance matrices. Based on the analysis in McCullagh and Nelder (1989), we will show that for a Poisson noise process, the MAP estimate is still equivalent to a weighted least squares fit, but with different weights. We have chosen the MAP estimate over the conditional mean (*i.e.*, minimum variance) or conditional median (*i.e.*, minimum absolute error) estimates because we must incorporate state constraints into the estimation process. This turns out to be quite easy if one is using weighted least squares, even

when the model is nonlinear.

It should be noted that extended Kalman filtering techniques are not appropriate for this tracking problem. First, the relationship between the solid angle subtended by the detector and the source location in Equation (2) is *not* analytic at the origin of the detector centered coordinate system. The resulting discontinuity in the first derivative makes any truncated Taylor series a poor approximation to the nonlinear sensor model. Second, the application of state constraints must be accomplished either by clipping the estimates or interactively adjusting the state covariance matrix. Both of these methods can lead to poor tracking performance. Third, extended Kalman filtering techniques assume a Gaussian noise distribution. Although a Poisson distribution can be approximated by a Gaussian distribution when the mean count rate is large, this approximation becomes less accurate as the count rate for a sensor decreases. Further details on the application and performance of extended Kalman filtering techniques for this problem are provided in Muske and Howse (2001).

We start our estimation process at time $k = 0$ with an estimate of the source position and background level $(x(0), y(0), z(0), \mathcal{B}(0))$. We assume that this initial estimate consists of the correct values at this time. We also assume that the absolute source strength \mathcal{S} is known, and that the correction factor product \mathcal{F}_i is constant over time. Note that these assumptions may not always be reasonable, but in our problem they are often valid. At succeeding time steps we update our estimate of the source location and background level using the relations

$$\begin{aligned} x(k) &= x(k-1) + \delta x^*(k), \\ y(k) &= y(k-1) + \delta y^*(k), \\ z(k) &= z(k-1) + \delta z^*(k), \\ \mathcal{B}(k) &= \mathcal{B}(k-1) + \delta \mathcal{B}^*(k), \end{aligned} \tag{4}$$

for all time steps $k = 1, 2, 3, \dots$, where $\delta x^*(k)$, $\delta y^*(k)$, $\delta z^*(k)$, and $\delta \mathcal{B}^*(k)$ are state changes which are optimal in some sense. Note that using our previous definitions, Equation (4) can be written in vector form as $\mathbf{e}(k) = \mathbf{e}(k-1) + \mathbf{d}^*(k)$.

At each time step k the optimal estimated state change $\mathbf{d}^*(k)$ is computed by solving the optimization problem

$$\begin{aligned} \mathbf{d}^*(k) &= \arg \min_{\mathbf{d}(k)} \left(\sum_{i=1}^m \mathcal{W}_i^m (\mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k)))^2 + \sum_{j=1}^p \mathcal{W}_j^d d_j^2(k) \right), \\ \text{subject to} \end{aligned} \tag{5}$$

$$\begin{aligned} -\mathcal{N}_j &\leq d_j(k) \leq \mathcal{N}_j, \\ \mathcal{L}_j &\leq e_j(k-1) + d_j(k) \leq \mathcal{U}_j, \quad j = 1, \dots, p, \quad k = 1, 2, \dots. \end{aligned}$$

In this equation, $\mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k))$ represents the mean value of our approximation of the detected count rate $\mathcal{E}\{\hat{\mathcal{D}}_i\}$, given by the right hand side of Equation (1) evaluated at $\mathbf{e}(k-1) + \mathbf{d}(k)$ for the i th detector. Note that evaluating the right hand side of Equation (1) requires solving Equations (2) and (3) first. The quantity \mathcal{W}_i^m is a weight which determines how closely the algorithm tries to match the model. Similarly, \mathcal{W}_j^d sets how much the algorithm is allowed to change the previous state estimate. The term $\sum_{i=1}^m \mathcal{W}_i^m (\mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k-1) + \mathbf{d}(k)))^2$ is a weighted L_2 norm of the difference $\|\mathcal{D}(k) - \mathcal{E}\{\hat{\mathcal{D}}(k)\}\|_2^2$ between the actual detected count rate $\mathcal{D}(k)$ and the expected value of the predicted detected count rate $\mathcal{E}\{\hat{\mathcal{D}}(k)\}$. This term corresponds to a weighted least squares fit of the state estimate to the measurements. Likewise

the term $\sum_{j=1}^p \mathcal{W}_j^d d_j^2(k)$ is the weighted L_2 norm of the difference $\|\mathbf{e}(k) - \mathbf{e}(k-1)\|_2^2$ between the estimated position and background at the previous time and that at the current time. This is a regularization term that penalizes large changes in the state estimate more than small changes. Equation (5) will try to minimize the change in position and background while simultaneously making a change that makes the resulting position and background approximately agree with the model. When selecting the weights, note that making \mathcal{W}_i^m large relative to \mathcal{W}_j^d reflects a high confidence in the sensor model, while setting \mathcal{W}_j^d high with respect to \mathcal{W}_i^m indicates confidence in the sensor measurements.

As stated previously, the MAP state estimate for a Poisson noise process is equivalent to a weighted least squares fit of the measurements with the appropriate weights. Maximizing the a posteriori probability $\mathcal{P}\{\mathbf{e}(k) | \mathcal{D}(k)\}$ is equivalent by Bayes rule to maximizing the probability $\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\} \mathcal{P}\{\mathbf{e}(k)\}$. Assuming that the measurements at a given time step $\mathcal{D}(k)$ are statistically independent, then given the source location $\mathbf{e}(k)$, the conditional probability for the measurements $\mathcal{D}(k)$ is simply the product of Poisson distributions

$$\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\} = \prod_{i=1}^m \exp\left(\mathcal{D}_i(k) \ln(\mathcal{M}_i(\mathbf{e}(k))) - \mathcal{M}_i(\mathbf{e}(k)) - \ln(\mathcal{D}_i(k)!) \right), \quad (6)$$

in which $\mathcal{M}_i(\mathbf{e}(k))$ is the mean count rate for the i th detector given the state $\mathbf{e}(k)$. Each component of this Poisson distribution has a variance equal to its mean, $\mathcal{V}(\mathcal{D}_i) = \mathcal{M}_i(\mathbf{e}(k))$. Assuming that all states $\mathbf{e}(k)$ are equally likely, we chose a uniform a priori probability $\mathcal{P}\{\mathbf{e}(k)\}$. The uniform a priori probability only scales the conditional probability $\mathcal{P}\{\mathcal{D}(k) | \mathbf{e}(k)\}$, and so does not affect the locations of extrema. Furthermore, the extremal points are invariant under the monotonic transformation $\ln(\cdot)$. Hence we will maximize the natural logarithm of the quantity in Equation (6), which we call the *likelihood function* $\mathcal{L}(\mathcal{D}; \mathcal{M})$. Maximizing $\mathcal{L}(\mathcal{D}; \mathcal{M})$ is equivalent to minimizing the difference between the largest possible value for $\mathcal{L}(\cdot)$, which is when $\mathcal{M} = \mathcal{D}$, and the value of $\mathcal{L}(\cdot)$ for the estimate \mathcal{M} . All of this discussion implies that the MAP estimate for $\mathbf{e}(k)$ under a Poisson distribution is obtained by minimizing the quantity

$$2 \sum_{i=1}^m \left(\mathcal{D}_i(k) \left(\ln \mathcal{D}_i(k) - \ln(\mathcal{M}_i(\mathbf{e}(k))) \right) - \left(\mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k)) \right) \right). \quad (7)$$

This result is analogous to the one obtained in McCullagh and Nelder (1989) for a univariate Poisson distribution.

It is stated in McCullagh and Nelder (1989) that many problems which ideally should be Poisson distributed, in practice have a value for the constant of proportionality between the variance and the mean which is greater than one. The constant of proportionality between the variance and the mean of such a Poisson-like distribution is called the *dispersion* and is denoted by γ_i^2 . For example, this phenomena occurs when samples of a Poisson process are made at random time intervals, rather than fixed intervals. This occurs in our problem because the length of time between count rate samples $\mathcal{D}_i(k)$ for each detector varies over time. Also, the order in which the different detectors are sampled varies with time. Furthermore, gamma detection involves conversion and amplification processes in the sensor which lead to a higher variance signal at the sensor output than was present at the sensor input, which results in $\gamma_i^2 > 1$. We assume that the dispersion is the same at each detector, making γ_i^2 independent of i , and denote this single quantity by γ^2 . We will discuss our choice for this dispersion parameter γ^2 in Section 4. The

MAP estimate for $\mathbf{e}(k)$ under a Poisson-like distribution where $\gamma^2 > 1$ is obtained by minimizing

$$2 \sum_{i=1}^m \frac{1}{\gamma^2} \left(\mathcal{D}_i(k) \left(\ln \mathcal{D}_i(k) - \ln(\mathcal{M}_i(\mathbf{e}(k))) \right) - \left(\mathcal{D}_i(k) - \mathcal{M}_i(\mathbf{e}(k)) \right) \right), \quad (8)$$

in analogy with Equation (7). An equivalence can be established between the *first* term in Equation (5) and Equation (8) by taking the first order Taylor series expansion of $\ln \mathcal{D}_i(k)$ about the point $\mathcal{M}_i(\mathbf{e}(k))$, and substituting this expansion into Equation (8). It is straightforward to show that the result of this substitution is identical to the first term in Equation (5) if the model weights are chosen as $\mathcal{W}_i^m = \frac{1}{\gamma^2 \mathcal{M}_i(\mathbf{e}(k))}$. Note that this choice of weights results in the diagonal weight matrix $\mathbf{W}^m = \text{diag}_{i=1,\dots,m}((\gamma^2 \mathcal{M}_i(\mathbf{e}(k)))^{-1})$.

The optimization problem posed in Equation (5) is a nonlinear programming problem with linear inequality constraints, which can be solved by a number of methods, many of which are discussed in Bertsekas (1999), Luenberger (1984) and Fletcher (1987). The optimization algorithm that we chose was CFSQP; the implementation is documented in Lawrence, Zhou, and Tits (1994) and the algorithm is discussed in Panier and Tits (1993) and Lawrence and Tits (1996). This algorithm uses a sequential quadratic programming (SQP) approach, modified so that each iteration is feasible with respect to the constraints. This algorithm is designed to minimize the maximum of a set of objective functions (*i.e.*, *minimax* problems), but since our set of objective functions contains only one member it also solves the problem posed in Equation (5). We chose this algorithm because all the intermediate iterates of the algorithm are feasible, so if we have to stop the optimization before achieving convergence, the resulting suboptimal solution will still satisfy the constraints. This is an important consideration in our application because we need to track the sources in real-time relative to the average detector sampling time, which in our case is 1 second. Hence if our optimization runs for more than 1 second at time step k we simply stop it, use the resulting suboptimal solution for the estimate $\mathbf{e}(k)$, and then begin estimating the next position $\mathbf{e}(k+1)$ with the new sensor data.

The method outlined so far computes the current state estimate $\mathbf{e}(k)$ based only on the current measurement $\mathbf{D}(k)$. In principle a better estimate may be obtained if the last l measurements $(\mathbf{D}(k), \mathbf{D}(k-1), \dots, \mathbf{D}(k-(l-1)))$ are used. This is the idea behind the moving horizon state estimation procedures described in Muske and Rawlings (1995) and Rao and Rawlings (2000). The optimal state change $\tilde{\mathbf{d}}^*(k)$ over the entire window is found by solving the optimization problem

$$\begin{aligned} \tilde{\mathbf{d}}^*(k) = \arg \min_{\tilde{\mathbf{d}}(k)} & \left(\sum_{n=0}^{l-1} \left(\sum_{i=1}^m \mathcal{W}_i^m \left(\mathcal{D}_i(k-n) - \mathcal{M}_i(\mathbf{e}(k-n-1) + \mathbf{d}(k-n)) \right)^2 + \right. \right. \\ & \left. \left. \sum_{j=1}^p \mathcal{W}_j^d d_j^2(k-n) \right) \right), \end{aligned} \quad (9)$$

subject to

$$\begin{aligned} -\mathcal{N}_j &\leq d_j(k-n) \leq \mathcal{N}_j, \\ \mathcal{L}_j &\leq e_j(k-n-1) + d_j(k-n) \leq \mathcal{U}_j, \quad j = 1, \dots, p, \quad n = 0, \dots, l-1, \quad k = 1, 2, \dots. \end{aligned}$$

This equation reduces to Equation (5) when the window length l is equal to 1. In Equation (9) the state change for the entire window at time step k is $\tilde{\mathbf{d}}(k) = [\mathbf{d}(k) \ \mathbf{d}(k-1) \ \dots \ \mathbf{d}(k-(l-1))]^\top$, and the state estimate is $\tilde{\mathbf{e}}(k) = [\mathbf{e}(k) \ \mathbf{e}(k-1) \ \dots \ \mathbf{e}(k-(l-1))]^\top$. This algorithm is initialized by performing a batch estimation until sample time l . Muske and Rawlings (1995) present examples

from several application areas in which this method outperforms one-step (*i.e.*, $l = 1$) recursive estimation.

4 Tracking Results

In this section we present tracking results based on experiments with a source in the facility. The experimental procedure consisted of placing a Cs¹³⁷ source on a cart and moving it along a predetermined path in the room. Along this path there were 17 points at which we paused with the source for approximately 3 minutes each. The purpose of the 3 minute pauses was to collect enough data at each position to compute a mean count rate for each sensor at each position. The resulting means are used to compare the performance of our algorithm on the actual data to its performance on idealized Poisson data. The estimated position of the source as seen from the ceiling looking down is shown in Figure 5 . Figure 5(a) shows the estimated source positions for

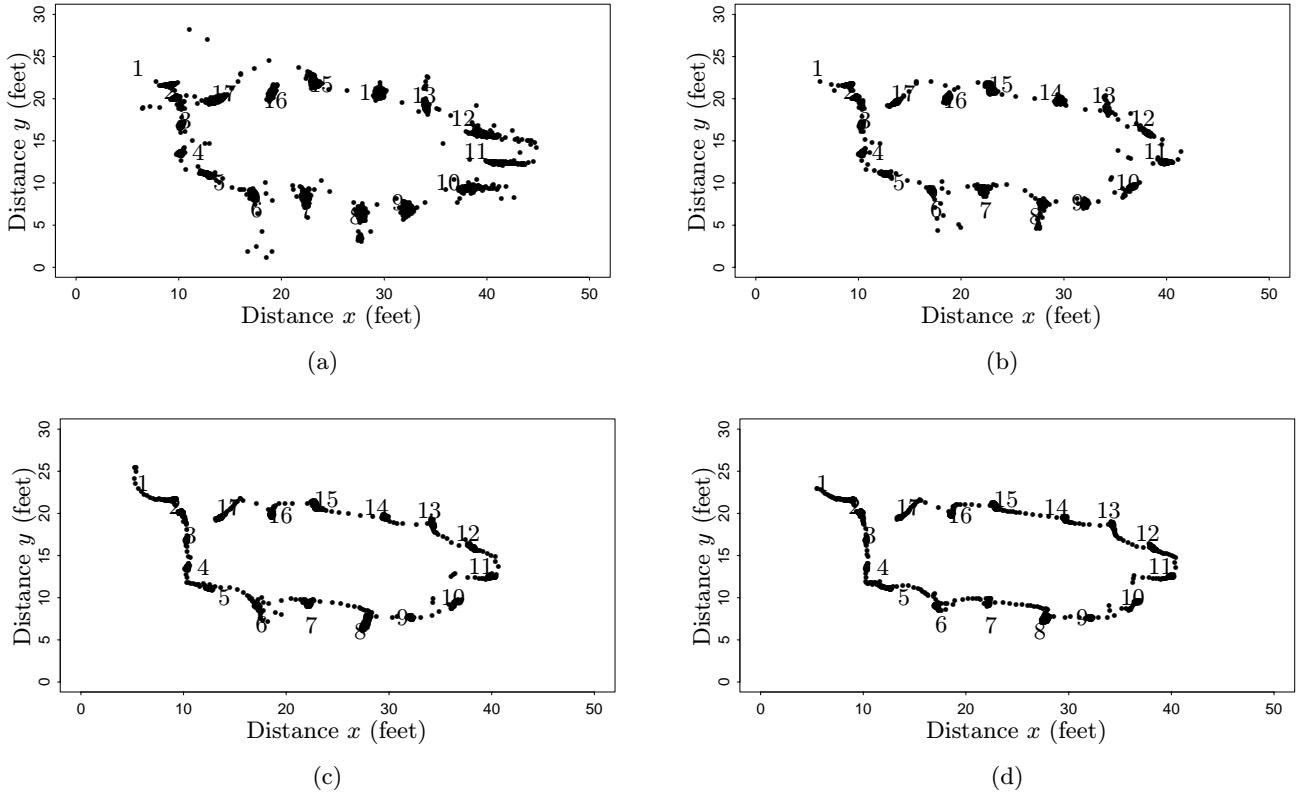


Figure 5: The estimated position of the source in the x - y plane with dispersion (a) $\gamma^2 = 1$ (b) $\gamma^2 = 10$ (c) $\gamma^2 = 50$ (d) $\gamma^2 = 100$ for window length $l = 1$. These plots show a view of the room from the ceiling looking down. The detectors are located in each of the four corners of these pictures.

dispersion $\gamma^2 = 1$, Figure 5(b) shows the locations with $\gamma^2 = 10$, Figure 5(c) illustrates positions for $\gamma^2 = 50$, and Figure 5(d) gives estimates for $\gamma^2 = 100$. In all cases the state weights are $W_j^d = 1$ and the window length is $l = 1$. The 17 numbers in these plots show the actual positions of the points at which we paused with the source. Beginning at point 1, we followed the numbered points in ascending order, ending at point 17. The points \bullet in the figure represent the pairs $(x(k), y(k))$

of the state estimate for every time step k . These estimates are computed at 1 second intervals, which is also the average sampling time for the detectors. Our algorithm runs fast enough for us to compute each state estimate in less than 1 second, hence we are able to track the source in real-time relative to the average sampling rate of the detector. Although the detectors do not appear in these pictures, they are located in each corner at a 45° angle with respect to the walls. Note that as the dispersion γ^2 is increased from Figure 5(a) to Figure 5(d), the position estimates go from clusters around the 17 numbered points, to a smooth track running through these 17 points.

Figure 5 shows that most of the estimates of the source location are close to the path described by the 17 points where we paused with the source. However, it does not show whether the estimates follow this path in the correct temporal order. Figure 6 shows the estimated positions in both the x and y directions versus time for the x - y plots shown in Figures 5(a) and 5(d). In

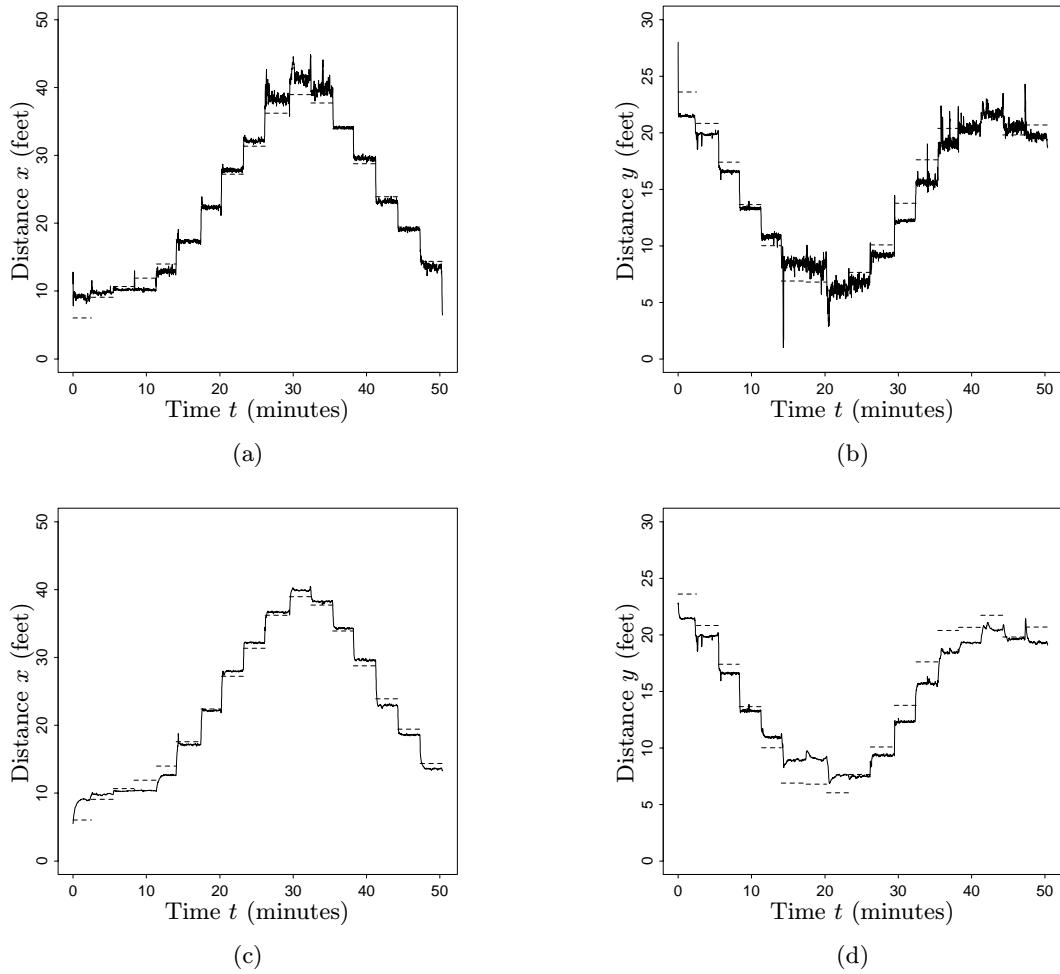


Figure 6: The estimated positions of the source for dispersion $\gamma^2 = 1$ in the (a) x -direction (b) y -direction versus time. The calculated source locations for dispersion $\gamma^2 = 100$ in the (c) x -direction (d) y -direction versus time. The dotted lines show the actual source positions over time.

these two plots the the solid lines represent the estimated source position, and the dotted lines represent the actual source position. Figures 6(a) and 6(c) plot the positions in the x -direction

versus time for dispersions $\gamma^2 = 1$ and $\gamma^2 = 100$ respectively. Likewise Figures 6(b) and 6(d) show the y -position with respect to time for the same two dispersions. It is clear from these pictures that the numbered points are visited in the correct order.

We found that removing both the regularization term $\|\mathbf{d}(k)\|_2^2$ and the constraints $-\mathcal{N} \leq \mathbf{d}(k) \leq \mathcal{N}$ and $\mathcal{L} \leq \mathbf{e}(k) \leq \mathcal{U}$ from Equation (5) resulted in the unregularized unconstrained algorithm failing to properly track the source. This modified algorithm tended to initially track the source with steadily increasing tracking error and then stall at some state value for which at least some of the states were infeasible. Recall that in our case the states x , y and z are infeasible if they are outside of the room and the state value \mathcal{B} is infeasible if it exceeds historical bounds for the background levels in the room. The rate at which the tracking error deteriorated tended to increase with increasing values of the dispersion γ^2 . Similarly the time step at which this modified algorithm finally stalled tended to decrease with increasing dispersion. When the regularization term $\|\mathbf{d}(k)\|_2^2$ was removed but the constraints were retained, the unregularized algorithm tracked the source fairly well for values of γ^2 near one, but the tracking error increased rather rapidly as γ^2 was increased. When the constraints $-\mathcal{N} \leq \mathbf{d}(k) \leq \mathcal{N}$ and $\mathcal{L} \leq \mathbf{e}(k) \leq \mathcal{U}$ were removed but the regularization term was retained, the unconstrained algorithm tracked the source almost as well as the original algorithm for values of γ^2 near one. For the unconstrained algorithm the tracking error increased as γ^2 was increased, but more slowly than for the unregularized algorithm. These results indicate that in order to use the dispersion γ^2 as a tuning parameter to adjust the tracking performance, both the regularization term and the constraints must be used.

We estimated the error of this tracking algorithm using the following two statistics. For both statistics we begin by taking the absolute value of the difference between the median estimated position and the actual position for each of the 17 positions in both the x and y directions. This corresponds to taking the median value of each plateau in Figure 6, subtracting each median from each dotted line, and taking the absolute value of each result. This results in two sets of 17 numbers, one for each direction x and y . For the first statistic, combine the elements within each set by taking the maximum value over the set. Call the resulting maximum absolute error in the x -direction \mathcal{E}_x^{max} , and in the y -direction call it \mathcal{E}_y^{max} . For the second statistic, take the median value to combine the differences within each set. Denote the median absolute error along x by \mathcal{E}_x^{med} and along y by \mathcal{E}_y^{med} . These two statistics are then used to compute two percent location error statistics for the algorithm. To find the maximum percent location error \mathcal{R}_e^{max} , construct a rectangle of width $2\mathcal{E}_x^{max}$ and height $2\mathcal{E}_y^{max}$. Compute the percent location error \mathcal{R}_e^{max} by dividing the area of this rectangle by the area of the room. Similarly the median percent location error \mathcal{R}_e^{med} is computed by dividing the area of the $2\mathcal{E}_x^{med} \times 2\mathcal{E}_y^{med}$ rectangle by the room area. For the dispersion values $\gamma^2 = 1$ and $\gamma^2 = 100$ shown in Figure 6, the values of these error statistics

γ^2	\mathcal{E}_x^{max} (feet)	\mathcal{E}_y^{max} (feet)	\mathcal{R}_e^{max} (%)	\mathcal{E}_x^{med} (feet)	\mathcal{E}_y^{med} (feet)	\mathcal{R}_e^{med} (%)
1	3.137	2.141	2.117	0.740	0.909	0.212
100	2.916	2.284	2.099	0.769	1.358	0.329

Table 1: Two error measures for the location estimates shown in Figure 6.

are given in Table 1. Based on these two error statistics it appears that the lower dispersion value $\gamma^2 = 1$ leads to a more accurate estimate than the higher dispersion value $\gamma^2 = 100$. Also it appears that increasing the dispersion increases the time lag between the estimated and actual locations.

The statistics in Table 1 become more meaningful when compared with the those from an idealized situation. As a baseline against which to evaluate our performance, we chose a scenario in which the gamma emissions from the source were assumed to be a Poisson process with dispersion $\gamma^2 = 1$, and the detectors were assumed to output the exact gamma count rate with no additional noise. To test the algorithm on this ideal scenario we generated synthetic data drawn from a Poisson process as the output of each detector. This synthetic experiment was identical to our actual experiment in that the synthetic source paused for the same length of time at the same 17 positions. The synthetic data was generated by first taking the average μ_l^i of the actual count rate for each detector $i = 1, \dots, 4$ at each position $l = 1, \dots, 17$. We then computed the number of measurements s_l made at each position $l = 1, \dots, 17$. Using these two numbers we drew s_l samples from a Poisson distribution with mean μ_l^i for all detectors $i = 1, \dots, 4$ and all positions $l = 1, \dots, 17$. Twenty-five synthetic data sets were created in this fashion, and our algorithm was run on each of them with a window length of $l = 1$. We then computed the median absolute error in both the x and y directions for the estimates generated with each of these twenty-five synthetic data sets. For comparison to the results with actual data, we calculated the mean and standard deviation of this statistic over the twenty-five synthetic runs. The results are presented in Table 2. Notice that neither \mathcal{E}_x^{med} and \mathcal{E}_y^{med} for the actual data fall within one standard deviation of the

	Mean \mathcal{E}_x^{med} (feet)	Deviation \mathcal{E}_x^{med} (feet)	Mean \mathcal{E}_y^{med} (feet)	Deviation \mathcal{E}_y^{med} (feet)
Actual	0.740	—	0.909	—
Synthetic	0.758	0.00439	0.922	0.00311

Table 2: A comparison of the median absolute error between the actual data and synthetic data drawn from a Poisson distribution.

means for the synthetic data.

We conducted a Wilcoxon test, described for instance in Conover (1980) and Hollander and Wolfe (1973), of the hypothesis that the value of \mathcal{E}_x^{med} for the actual data was drawn from a distribution having the same mean as the one for the synthetic data. We conducted a similar hypothesis test for the statistic \mathcal{E}_y^{med} . Note that this pair of tests considers the x and y coordinates separately. Therefore each hypothesis test is on a univariate distribution and correlation between x and y is *not* considered. These hypothesis tests gave p -values of 0.0006 and 0.0003 for \mathcal{E}_x^{med} and \mathcal{E}_y^{med} respectively. We also conducted a Hotelling t^2 -test, described in Johnson and Wichern (1992), of the hypothesis that the value $(\mathcal{E}_x^{med}, \mathcal{E}_y^{med})$ for the actual data was drawn from a distribution having the same mean as the one for the synthetic data. Note that this test considers the x and y coordinates together. Therefore the hypothesis test is on a bivariate distribution and correlation between x and y is considered. This hypothesis test gave a p -value of 0.0003. The p -value specifies the *maximum* significance at which the null hypothesis should be accepted. So if the desired level of significance is less than the p -value, accept the null hypothesis, otherwise reject it. The level of significance is the probability of accepting the alternative hypothesis when the null hypothesis is true, so it is conceptually identical to a maximum false alarm rate. We chose 0.01 (*i.e.*, a 1% false alarm probability) as a reasonable significance, therefore with 99% confidence \mathcal{E}_x^{med} and \mathcal{E}_y^{med} for the actual data came from a distribution whose mean is *different* from the synthetic data distribution. Note that the closer the p -value is to one, the more likely it is that the mean of the actual distribution is equal to the mean of the synthetic distribution. Several conclusions emerge from this analysis. First, the performance of our algorithm on real data is qualitatively very similar to its performance on idealized synthetic data, which indicates

that our model captures the major features of the real data. Second, since the errors for the actual data have a different mean than those for the synthetic data, our assumption that the dispersion γ^2 is one is probably incorrect. Third, since the median absolute errors for both the actual and synthetic data are not zero, there is probably a minor systematic error in our sensor model. We conjecture that the primary source of unmodeled error comes from neglecting the change in gamma absorption with position due to the finite detector thickness, but we have not attempted to model this phenomena.

We now address the issue of choosing an appropriate value for the dispersion. It is clear from Figures 5 and 6 that the deviation between the estimated and actual position, called the *bias*, changes as the dispersion γ^2 changes. The estimator bias is given mathematically by the expression $\mathcal{B} = |\mathcal{E}\{\mathbf{e}\} - \mathcal{E}\{\mathbf{p}\}|$, where $\mathcal{E}\{\mathbf{p}\}$ is the true source position and background count rate. Note that this definition means that our error statistics from the previous paragraph are the maximum model bias and the median model bias as a percentage of the room size. It can also be seen in Figures 5 and 6 that the deviation between the estimated position and the average estimated position, called the *variance*, also changes with dispersion. The estimator variance is given by $\mathcal{V} = \mathcal{E}\{(\mathbf{e} - \mathcal{E}\{\mathbf{e}\})^2\}$. One way to choose the dispersion parameter γ^2 is by selecting the value which minimizes the sum of the squared bias and the variance. Since the sum of squared bias and variance represents the total error made by the model, this choice of dispersion will minimize the total error. Plots of bias, standard deviation, and total error for the x - y plane for a range of dispersion values are shown in Figure 7. Figure 7(a) plots the estimator bias versus dispersion γ^2 ,

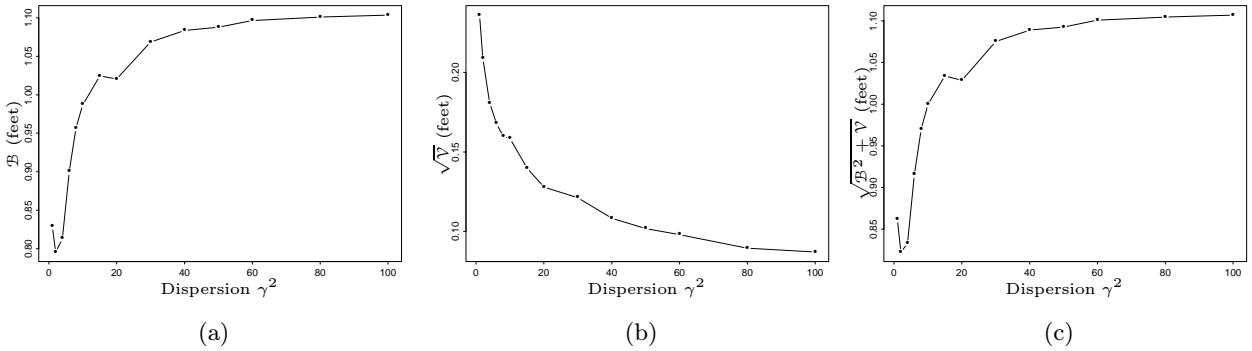


Figure 7: The change in (a) estimator bias (b) estimator standard deviation (c) total estimator error as a function of the dispersion parameter γ^2 .

Figure 7(b) shows the change in standard deviation with dispersion, and Figure 7(c) illustrates the variation in total estimator error as dispersion changes. Comparing the relative sizes of the errors in Figures 7(a) and 7(b), it is clear that bias is the dominant error source in this problem. Also note that bias generally increases as dispersion increases, while standard deviation decreases with increasing dispersion. This seems reasonable if the dispersion is considered a smoothness parameter, since increasing the smoothness tends to decrease standard deviation and increase bias. From Figure 7(c) it appears that the minimum value for the total error occurs at the dispersion $\gamma^2 = 2$. Analyzing the variance to mean ratio of the data gives an average dispersion value of 1.60, which is in excellent agreement with Figure 7(c).

We found that with the system model shown in Figure 4, the state estimates for all window lengths $l > 1$ were almost identical to the estimates for window length $l = 1$. We believe that this observation can be explained by the fact that the model in Figure 4 has no dynamics, hence any change in the estimate $\mathbf{e}(k-1)$ is completely determined by the external driving force $\mathbf{d}(k)$. While

it may appear that the regularization term $\|\mathbf{e}(k) - \mathbf{e}(k-1)\|_2^2$ provides some coupling between the states at different times, recall that the current state estimate $\mathbf{e}(k)$ is selected so as to minimize this term. Therefore the observation that there is only very weak coupling between states at different times seems reasonable. If the model contained terms approximating the source velocities, then perhaps a moving horizon state estimate with window length $l > 1$ would be different from a state estimate with window length $l = 1$. However, velocity estimation has questionable value in this problem because a person can change velocities much more quickly than our average detector sampling rate.

5 Conclusion

We have described a real-time algorithm for tracking the position of radioactive sources in a facility in the presence of measurement noise. The tracking algorithm was formulated as a nonlinear state estimation problem. A constrained, feasible path, sequential quadratic program was used to solve the resulting nonlinear, recursive least squares optimization problem. The optimization simultaneously minimizes the change in source position and disagreement between measurements and a sensor model. The sensor model is a fairly complex nonlinear function relating position to detected count rate.

The overall purpose of this work is to enhance both security and safety by automating part of the assessment process, allowing remote assessment, and introducing new sensor modalities into the assessment process. We presented tracking results based on an experiment done with one source in a single room. Our results indicate that a source can be tracked quite well with this algorithm in spite of rather high measurement noise levels. We have demonstrated the capability to track a single source in real-time with high accuracy in spite of a nonlinear sensor model and high measurement noise.

Acknowledgments

We would like to thank Kevin Buescher and Don Hush for valuable insights and enlightening discussions in the course of this work.

References

- Bertsekas, D. (1999). *Nonlinear Programming* (2nd edition). Optimization and Neural Computation Series. Athena Scientific, Belmont, MA.
- Bryson, A., & Ho, Y.-C. (1975). *Applied Optimal Control: Optimization, Estimation, and Control* (Revised edition). Hemisphere Publishing Corp., Bristol, PA.
- Conover, W. (1980). *Practical Nonparametric Statistics* (2nd edition). Applied Probability and Statistics Series. John Wiley & Sons, Inc., New York, NY.
- Fletcher, R. (1987). *Practical Methods of Optimization* (2nd edition). John Wiley & Sons, Ltd., Chichester, United Kingdom.
- Gotoh, H., & Yagi, H. (1971). Solid Angle Subtended by a Rectangular Slit. *Nuclear Instruments and Methods*, 96(2), 485–486.

- Hollander, M., & Wolfe, D. (1973). *Nonparametric Statistical Methods*. Applied Probability and Statistics Series. John Wiley & Sons, Inc., New York, NY.
- Jazwinski, A. (1970). *Stochastic Processes and Filter Theory*, Vol. 64 of *Mathematics in Science and Engineering*. Academic Press, Inc., New York, NY.
- Johnson, R., & Wichern, D. (1992). *Applied Multivariate Statistical Analysis* (3rd edition). Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Lawrence, C., & Tits, A. (1996). Nonlinear Equality Constraints in Feasible Sequential Quadratic Programming. *Optimization Methods and Software*, 6(2), 265–282.
- Lawrence, C., Zhou, J., & Tits, A. (1994). User's Guide for CFSQP. Tech. rep. TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, MD.
- Luenberger, D. (1984). *Linear and Nonlinear Programming* (2nd edition). Addison-Wesley Publishing Co., Inc., Reading, MA.
- McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models* (2nd edition),, Vol. 37 of *Statistics and Applied Probability*. Chapman & Hall, Ltd., London, United Kingdom.
- Muske, K., & Howse, J. (2001). Comparison of Recursive Estimation Techniques for Position Tracking Radioactive Sources. In *Proceedings of the 2001 American Control Conference*. American Automatic Control Council. To Appear.
- Muske, K., & Rawlings, J. (1995). Nonlinear Moving Horizon State Estimation. In Berber, R. (Ed.), *Methods of Model Based Process Control*, pp. 349–365. Kluwer Academic Publishers, Dordrecht, Netherlands.
- Panier, E., & Tits, A. (1993). On Combining Feasibility, Descent and Superlinear Convergence in Inequality Constrained Optimization. *Mathematical Programming*, 59(2), 261–276.
- Rao, C., & Rawlings, J. (2000). Nonlinear Moving Horizon Estimation. In Allgöwer, F., & Zheng, A. (Eds.), *Nonlinear Model Predictive Control*, Vol. 26 of *Progress in Systems and Control Theory*, pp. 45–69. Birkhäuser, Cambridge, MA.
- Sage, A., & Melsa, J. (1971). *Estimation Theory with Applications to Communications and Control*. Systems Science Series. McGraw-Hill, Inc., New York, NY.
- Tsoulfanidis, N. (1983). *Measurement and Detection of Radiation*. Nuclear Engineering Series. McGraw-Hill, Inc., New York, NY.